MATEMÁTICA APLICADA À COMPUTAÇÃO

AULA 5

Prof. Ricardo Alexandre Deckmann Zanardini

CONVERSA INICIAL

Chegou o momento de estudarmos temas relacionados à estatística e à probabilidade. Iniciaremos com conceitos e exemplos relacionados a média, moda e mediana. Aprenderemos a obter esses valores de forma analítica e também por meio do Python. Em seguida, estudaremos variância, desvio padrão, conceitos de probabilidade e de distribuição normal. Para finalizarmos, abordaremos técnicas de amostragem e intervalo de confiança.

TEMA 1 – ESTATÍSTICA DESCRITIVA (MÉDIA, MEDIANA, MODA)

Quando estamos trabalhando com um conjunto de dados, é muito comum precisarmos de valores que nos tragam uma ideia de valores associados a esses dados. Um desses valores é a média. É muito comum pensarmos em um salário médio de uma profissão ou o preço médio de uma determinada mercadoria.

A média é um número que representa a tendência central de um conjunto de dados. Representamos a média por $\bar{X}$ e a obtemos pela soma dos valores de um conjunto de dados x_i dividida pelo total de elementos:

$$\bar{X} = \frac{\sum x_i}{n}$$

Exemplo:

Uma indústria alimentícia produz *cookies* americanos. No processo de fabricação, cada porção de massa é pesada antes de ser assada para que os *cookies* sejam produzidos dentro do mesmo padrão. Espera-se que cada porção tenha 50 gramas, mas por diversos fatores há uma variação no peso das porções. A seguir, temos uma amostra contendo os pesos, em gramas, de 20 porções de massa:

49,7, 50,9, 48,9, 49,8, 50,1, 50,2, 50,8, 49,2, 50,1, 50,0,

50,4, 48,8, 49,3, 49,5, 49,1, 50,6, 49,0, 49,7, 49,7, 50,2

Com base nesses dados, obtenha o respectivo peso médio.

Como a média é obtida pela divisão da soma dos dados pelo número total de elementos, temos:

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\sum x_i = 49{,}7 + 50{,}9 + 48{,}9 + 49{,}8 + 50{,}1 + 50{,}2 + 50{,}8 + 49{,}2 + 50{,}1 + 50{,}0$$
$$+ 50{,}4 + 48{,}8 + 49{,}3 + 49{,}5 + 49{,}1 + 50{,}6 + 49{,}0 + 49{,}7 + 49{,}7$$
$$+ 50{,}2 = 996$$

$$\bar{X} = \frac{996}{20}$$

$$\bar{X} = 49{,}8$$

O peso médio corresponde a 49,8 g.

Além da média, outra medida de posição bastante utilizada é a moda. A moda, indicada por Mo, está associada ao dado que aparece o maior número de vezes em determinado conjunto. Podemos ter mais do que um único elemento relacionado à moda. Se não houver dados que se repetem, não há moda, e o conjunto é chamado de amodal.

Exemplo:

Uma indústria alimentícia produz *cookies* americanos. No processo de fabricação, cada porção de massa é pesada antes de ser assada para que os *cookies* sejam produzidos dentro do mesmo padrão. Espera-se que cada porção tenha 50 gramas, mas por diversos fatores há uma variação no peso das porções. A seguir, temos uma amostra contendo os pesos, em gramas, de 20 porções de massa:

49,7, 50,9, 48,9, 49,8, 50,1, 50,2, 50,8, 49,2, 50,1, 50,0,

50,4, 48,8, 49,3, 49,5, 49,1, 50,6, 49,0, 49,7, 49,7, 50,2

Obtenha a respectiva moda.

Para facilitar o processo, inicialmente, vamos ordenar o conjunto de dados:

48,8, 48,9, 49,0, 49,1, 49,2, 49,3, 49,5, **49,7**, **49,7**, **49,7**,

49,8, 50,0, 50,1, 50,1, 50,2, 50,2, 50,4, 50,6, 50,8, 50,9

Podemos identificar que o valor com a maior frequência, ou seja, o valor que aparece o maior número de vezes é 49,7. Logo, Mo = 49,7.

Temos também outra medida de posição, a mediana, representada por Md. A mediana corresponde ao elemento que ocupa a posição central de um rol, ou seja, o elemento que está no centro de um conjunto de dados ordenados. Quando o número total de dados é ímpar, há um único elemento ocupando a posição central. No caso de um número par de dados, dois elementos ocupam a posição central do rol. Quando isto ocorre, a mediana é obtida pela média entre esses dois elementos.

Como a mediana divide o conjunto de dados em duas partes iguais, aproximadamente metade dos dados está abaixo da mediana, e aproximadamente metade dos dados está acima da mediana.

Exemplo:

Uma indústria alimentícia produz *cookies* americanos. No processo de fabricação, cada porção de massa é pesada antes de ser assada para que os *cookies* sejam produzidos dentro do mesmo padrão. Espera-se que cada porção tenha 50 gramas, mas por diversos fatores há uma variação no peso das porções. A seguir, temos uma amostra contendo os pesos, em gramas, de 20 porções de massa:

49,7, 50,9, 48,9, 49,8, 50,1, 50,2, 50,8, 49,2, 50,1, 50,0,

50,4, 48,8, 49,3, 49,5, 49,1, 50,6, 49,0, 49,7, 49,7, 50,2

Com base nesses dados, obtenha a mediana.

Colocando os dados em ordem crescente, temos:

48,8, 48,9, 49,0, 49,1, 49,2, 49,3, 49,5, 49,7, 49,7, 49,7,

49,8, 50,0, 50,1, 50,1, 50,2, 50,2, 50,4, 50,6, 50,8, 50,9

Para identificarmos qual é a posição central, vamos dividir o total de elementos por 2:

$$\frac{20}{2} = 10$$

O número de elementos é par. Como o resultado da divisão é igual a 10, os elementos que ocupam a 10° e a 11° posição estão no centro do rol. Esses valores são 49,7 e 49,8. Portanto, a mediana é dada por:

$$Md = \frac{49{,}7 + 49{,}8}{2}$$

$$Md = \frac{99{,}5}{2}$$

$$Md = 49{,}75$$

A mediana, neste caso, é igual a 49,75. Podemos observar que os valores 48,8, 48,9, 49,0, 49,1, 49,2, 49,3, 49,5, 49,7, 49,7 e 49,7 estão abaixo da mediana e que os valores 49,8, 50,0, 50,1, 50,1, 50,2, 50,2, 50,4, 50,6, 50,8 e 50,9 estão acima da mediana.

Bibliotecas do Python foram desenvolvidas para cálculos estatísticos. Utilizaremos alguns recursos da biblioteca Pandas, desenvolvida para a análise e o tratamento de dados. Podemos inserir os dados manualmente ou importar os dados de bases armazenadas em planilhas, *sites*, bancos de dados, entre outros. Aprenderemos a inserir dados e também a importar dados do Excel. Com base nesses dados, neste momento, vamos obter a média, a moda e a mediana.

Exemplo:

Uma indústria alimentícia produz *cookies* americanos. No processo de fabricação, cada porção de massa é pesada antes de ser assada para que os *cookies* sejam produzidos dentro do mesmo padrão. Espera-se que cada porção tenha 50 gramas, mas por diversos fatores há uma variação no peso das porções. A seguir, temos uma amostra contendo os pesos, em gramas, de 20 porções de massa:

49,7, 50,9, 48,9, 49,8, 50,1, 50,2, 50,8, 49,2, 50,1, 50,0,

50,4, 48,8, 49,3, 49,5, 49,1, 50,6, 49,0, 49,7, 49,7, 50,2

Com base nesses dados, por meio do Python, obtenha a média, a moda e a mediana.

Precisamos importar a biblioteca Pandas e armazenar os dados na variável *x*. Como os dados equivalem a uma coluna de uma planilha, atribuímos o nome *Pesos* a esses dados. Em seguida, transformaremos os dados em um *Data Frame*, ou seja, uma matriz com um nome associado a cada coluna. Finalmente, podemos obter a média, a moda e a mediana por meio dos comandos *mean*, *mode* e *median*, respectivamente. Utilizaremos f-string para apresentar os resultados. A sequência completa dos comandos e as respectivas sintaxes são:

Figura 1 – Sequência completa dos comandos e respectivas sintaxes

```
import pandas as pd

x={'Pesos':[48.8, 48.9, 49.0, 49.1, 49.2, 49.3, 49.5, 49.7, 49.7, 49.7, 49.8, 50.0, 50.1, 50.1, 50.2, 50.2, 50.4, 50.6, 50.8, 50.9]}
```

```
p=pd.DataFrame(x)

media=p['Pesos'].mean()

moda=p['Pesos'].mode()

mediana=p['Pesos'].median()

print(f'Média: {media}')

print(f'Moda: {moda}')

print(f'Mediana: {mediana}')
```

Os resultados obtidos são:

Figura 2 – Resultados obtidos

```
Média: 49.8

Moda: 0    49.7

dtype: float64

Mediana: 49.75
```

A média corresponde a 49,8, e a mediana corresponde a 49,75. Em relação à moda, o resultado 49,7 está associado à primeira posição (posição de índice 0) de um *array*. Isto ocorre pois podemos ter mais do que um valor associado à moda.

Para resolvermos o mesmo problema, mas agora importando os dados de uma planilha, o primeiro passo é digitarmos os dados do problema em uma planilha do Excel. É importante adicionar um título à coluna para que possamos identificar em qual coluna os dados estão contidos.

Quadro 1 – Dados do problema digitados em uma coluna

	A
1	Pesos
2	49,7
3	50,9

4	48,9
5	49,8
6	50,1
7	50,2
8	50,8
9	49,2
10	50,1
11	50,0
12	50,4
13	48,8
14	49,3
15	49,5
16	49,1
17	50,6
18	49,0
19	49,7
20	49,7
21	50,2

Agora, precisamos importar duas bibliotecas: a Pandas para realizar o cálculo da média, da moda e da mediana e a Io para a importação da planilha. Precisamos também importar a função *files* do google.colab. A planilha com os dados que está salva no computador é importada pelo comando uploaded=files.upload() que gera um botão para selecionar a planilha e fazer o respectivo *upload*. A partir dos comandos

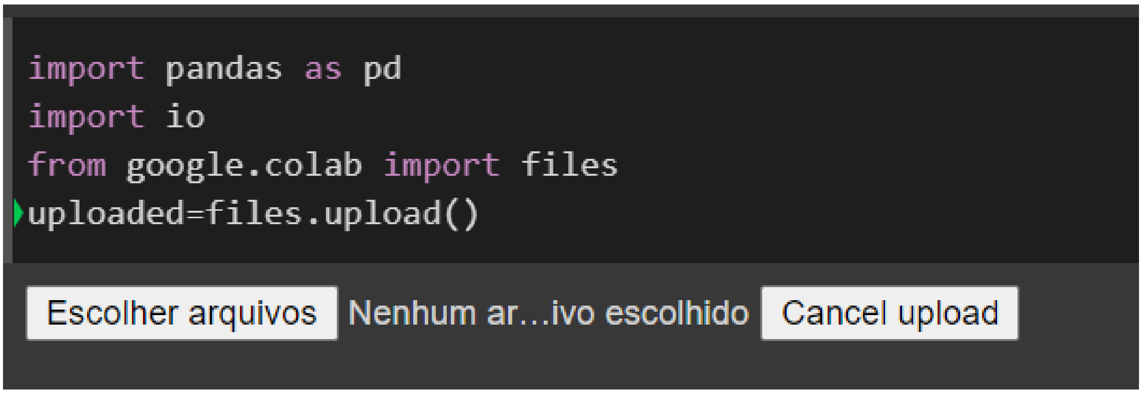
Figura 3 – Comandos

```
import pandas as pd

import io

from google.colab import files
```

```
uploaded=files.upload()
```

Temos:

Figura 4 – Resultado dos comandos



Clicando em “Escolher arquivos” e selecionando a planilha desejada, temos uma mensagem que informa que a planilha foi importada. A sequência a seguir armazena em *p* os dados da planilha, calcula a média, a moda e a mediana e apresenta os resultados obtidos:

Figura 5 – Sequência de comandos

```
p=pd.read_excel(io.BytesIO(uploaded['ExemploAula05.xlsx']))

media=p['Pesos'].mean()

moda=p['Pesos'].mode()

mediana=p['Pesos'].median()

print(f'Média: {media}')

print(f'Moda: {moda}')

print(f'Mediana: {mediana}')
```

Os respectivos resultados são:

Figura 6 – Resultados da sequência

```
Média: 49.80000000000001

Moda: 0     49.7
```

```
dtype: float64

Mediana: 49.75
```

TEMA 2 – ESTATÍSTICA DESCRITIVA (VARIÂNCIA, DESVIO PADRÃO)

Além de números que nos apresentam valores associados a um conjunto de dados, tais como a média, a moda e a mediana, muitas vezes precisamos de informações relacionadas à dispersão dos dados, ou seja, se esses dados estão muito afastados em relação à média ou mais concentrados. Essas medidas são chamadas de *medidas de dispersão*. Estudaremos o desvio padrão, que indica uma faixa de afastamento em relação à média que contém a maior parte dos dados. Como o desvio padrão é a raiz quadrada da variância, primeiro vamos estudar a variância.

Representamos a variância por σ^2, que é a média dos quadrados dos desvios de cada dado em relação à média. Há duas fórmulas para a variância, uma quando estamos trabalhando com dados relacionados à população e outra quando estamos trabalhando com uma amostra dos dados ou com uma população pequena, em que o total de termos é menor do que 30 valores. A variância populacional é obtida por meio da fórmula

$$\sigma^2 = \frac{\sum(x_i - \bar{X})^2}{n}$$

e a variância amostral é dada por

$$\sigma^2 = \frac{\sum(x_i - \bar{X})^2}{n-1}$$

A variância é um importante elemento para o cálculo do desvio padrão.

Exemplo:

Uma indústria alimentícia produz *cookies* americanos. No processo de fabricação, cada porção de massa é pesada antes de ser assada para que os *cookies* sejam produzidos dentro do mesmo padrão. Espera-se que cada porção tenha 50 gramas, mas por diversos fatores há uma variação no peso das porções. A seguir, temos uma amostra contendo os pesos, em gramas, de 20 porções de massa:

49,7, 50,9, 48,9, 49,8, 50,1, 50,2, 50,8, 49,2, 50,1, 50,0,

50,4, 48,8, 49,3, 49,5, 49,1, 50,6, 49,0, 49,7, 49,7, 50,2

Calcule a variância amostral.

Para obtermos a variância, primeiro precisamos da média:

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\sum x_i = 49{,}7 + 50{,}9 + 48{,}9 + 49{,}8 + 50{,}1 + 50{,}2 + 50{,}8 + 49{,}2 + 50{,}1 + 50{,}0 + 50{,}4 + 48{,}8 + 49{,}3 + 49{,}5 + 49{,}1 + 50{,}6 + 49{,}0 + 49{,}7 + 49{,}7 + 50{,}2 = 996$$

$$\bar{X} = \frac{996}{20}$$

$$\bar{X} = 49{,}8$$

Agora que temos o peso médio, podemos calcular $(x_i - \bar{X})^2$ e somar os resultados obtidos. Podemos realizar os cálculos com os dados ordenados ou não. Para facilitar a visualização, vamos trabalhar com os dados ordenados.

Tabela 1 – Dados para cálculo da variância

x_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
48,8	-1,0	1
48,9	-0,9	0,81
49,0	-0,8	0,64
49,1	-0,7	0,49
49,2	-0,6	0,36
49,3	-0,5	0,25
49,5	-0,3	0,09
49,7	-0,1	0,01
49,7	-0,1	0,01
49,7	-0,1	0,01
49,8	0,0	0
50,0	0,2	0,04
50,1	0,3	0,09
50,1	0,3	0,09

50,2	0,4	0,16
50,2	0,4	0,16
50,4	0,6	0,36
50,6	0,8	0,64
50,8	1,0	1
50,9	1,1	1,21
Total:	---	7,42

Podemos utilizar a fórmula

$$\sigma^2 = \frac{\sum(x_i - \bar{X})^2}{n-1}$$

Logo,

$$\sigma^2 = \frac{7,42}{20-1}$$

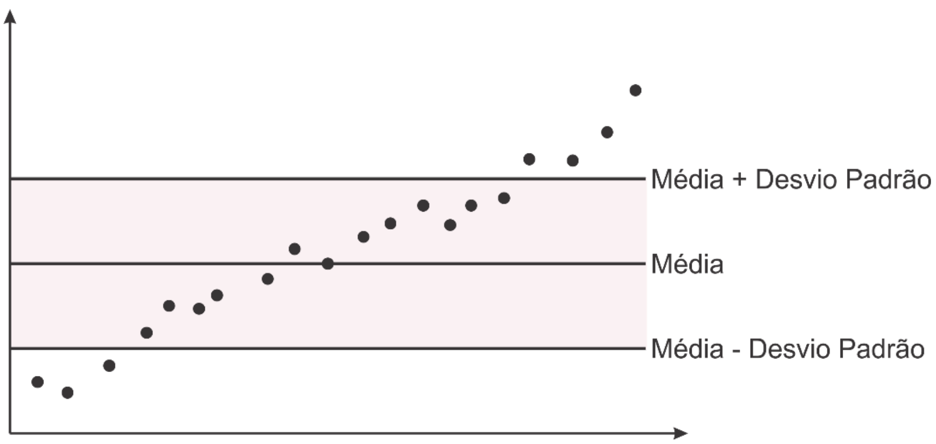
$$\sigma^2 = \frac{7,42}{19}$$

$$\sigma^2 = 0,39$$

A variância corresponde a 0,39.

Vamos agora calcular a raiz quadrada da variância, que é o desvio padrão. O desvio padrão indica a amplitude de uma faixa com centro na média que concentra a maior parte dos dados, geralmente 65% a 80% dos dados. Quanto maior o desvio padrão, maior a dispersão dos dados, e quanto menor o desvio padrão, menor a dispersão.

Gráfico 1 – Desvio padrão



O desvio padrão populacional é calculado por meio da fórmula

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n}}$$

e o desvio padrão amostral é calculado por meio da fórmula

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}}$$

Considerando o exemplo da indústria alimentícia que gerou uma variância igual a 0,39, podemos calcular o desvio padrão amostral fazendo

$$\sigma = \sqrt{0{,}39}$$

$$\sigma = 0{,}62$$

Podemos concluir que, com base na média 49,8, temos uma faixa de amplitude 0,62 acima da média e 0,62 abaixo da média que contém a maior parte dos dados.

Podemos utilizar o Python para calcularmos o desvio padrão, tanto da amostra quanto da população, de um conjunto de dados. O desvio padrão pode ser obtido por meio do comando std().

Vamos resolver um exemplo para aprendermos a utilizar o Python no cálculo do desvio padrão.

Exemplo:

Uma indústria alimentícia produz *cookies* americanos. No processo de fabricação, cada porção de massa é pesada antes de ser assada para que os *cookies* sejam produzidos dentro do mesmo padrão. Espera-se que cada porção tenha 50 gramas, mas por diversos fatores há uma variação no peso das porções. A seguir, temos uma amostra contendo os pesos, em gramas, de 20 porções de massa:

49,7, 50,9, 48,9, 49,8, 50,1, 50,2, 50,8, 49,2, 50,1, 50,0,

50,4, 48,8, 49,3, 49,5, 49,1, 50,6, 49,0, 49,7, 49,7, 50,2

Com base nesses dados, obtenha o desvio padrão por meio do Python.

O processo segue o mesmo princípio realizado para o cálculo da média. O primeiro passo é importar a biblioteca Pandas e armazenar os dados em uma variável *x*, atribuindo o nome *Pesos* aos dados. Depois, precisamos transformar os dados em um *Data Frame* e utilizar o comando std() para o cálculo do desvio padrão. A sequência de comandos é:

Figura 7 – Comandos

```
import pandas as pd

x={'Pesos':[48.8, 48.9, 49.0, 49.1, 49.2, 49.3, 49.5, 49.7,
49.7, 49.7, 49.8, 50.0, 50.1, 50.1, 50.2, 50.2, 50.4, 50.6,
50.8, 50.9]}

p=pd.DataFrame(x)

desviopadrao=p['Pesos'].std()

print(f'Desvio padrão: {desviopadrao}')
```

que resulta em:

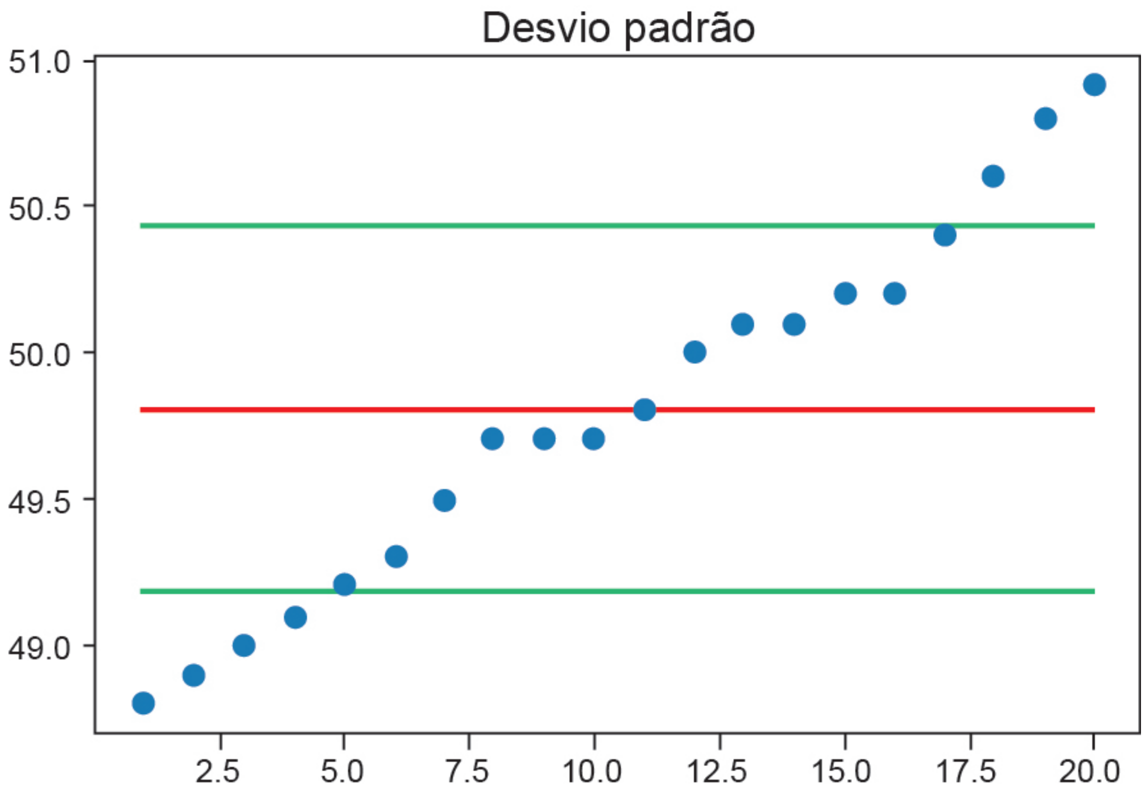
Figura 8 – Resultado dos comandos

```
Desvio padrão: 0.6249210476447995
```

Logo, o desvio padrão é igual a 0,62.

Graficamente, a média é representada pela linha horizontal vermelha. As linhas verdes indicam o valor da média menos o desvio padrão e o valor da média mais o desvio padrão. Observe que, dos 20 valores, 13 deles estão nessa faixa, o que corresponde a 65% dos dados.

Gráfico 2 – Desvio padrão



TEMA 3 – PROBABILIDADE (CONCEITOS DE PROBABILIDADE, DISTRIBUIÇÃO NORMAL)

Quando pensamos em probabilidade, é comum pensarmos na chance de ganharmos determinado jogo, na chance de chover ou não, na chance de obtermos um bom resultado em uma avaliação. De um modo geral, a probabilidade está associada à análise de situações em que há possibilidades diferentes de acontecimentos. Por meio de métodos existentes, é possível obter medidas que indicam quais são as chances de que determinado evento ocorra.

Dizemos que a probabilidade corresponde à medida de ocorrência de determinado evento aleatório que ocorre em um espaço amostral, conjunto de todos os possíveis resultados obtidos de um experimento estatístico e um evento aleatório.

Com base em dados ou em observações, podemos ter determinado grau de confiança, ou seja, uma probabilidade maior ou menor de ocorrência de determinado evento.

Figura 9 – Probabilidade de ocorrência de um evento

A probabilidade *P* de ocorrência de um fato *A*, chamada de P(A), corresponde à divisão do número de elementos *n* do evento *A* pelo número *N* de elementos do espaço amostral:

$$P(A) = \frac{n}{N}$$

Exemplo:

Em uma produção de 1000 peças, 14 delas apresentam defeitos que impossibilitam o uso. Qual é a probabilidade de que uma dessas peças, escolhida aleatoriamente, não apresente defeito?

Como o número de peças defeituosas é igual a 14 e foram produzidas 1000 unidades, temos 1000- 1 4 = 986 peças não defeituosas. Logo, n = 986 e N = 1000. A probabilidade de que uma peça não defeituosa seja escolhida é calculada por meio da fórmula

$$P(A) = \frac{n}{N}$$

Sendo assim, temos

$$P(A) = \frac{986}{1000}$$

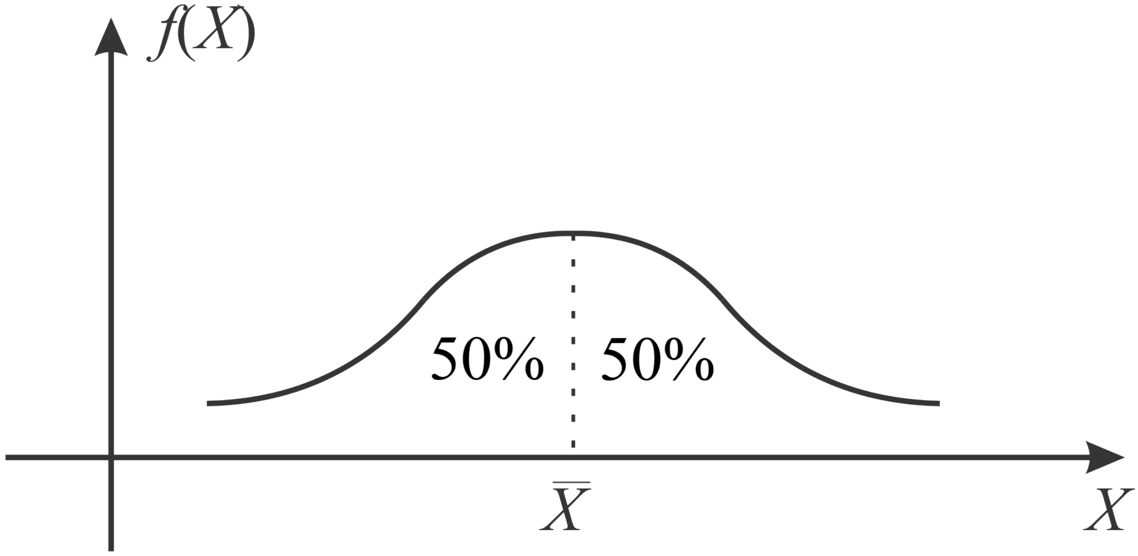
$$P(A) = 0{,}986$$

$$P(A) = 98{,}6\%$$

Portanto, a probabilidade de que uma peça não defeituosa seja escolhida é igual a 0,986 ou, de forma equivalente, 98,6%.

Além da probabilidade de ocorrência de determinado evento, podemos pensar em distribuição de probabilidade, ou seja, na descrição de probabilidade associada a um valor da variável *X* ou da probabilidade de que a variável *X* esteja em determinado intervalo. Dentre diferentes distribuições de probabilidade, uma delas é a distribuição normal. Uma distribuição normal tem a seguinte característica: apresenta uma concentração maior de dados próximos da média e uma quantidade menor de dados afastados da média. Graficamente, uma distribuição normal possui um formato parecido com um sino e é simétrica em relação à média. Essa curva é chamada de *curva de Gauss*.

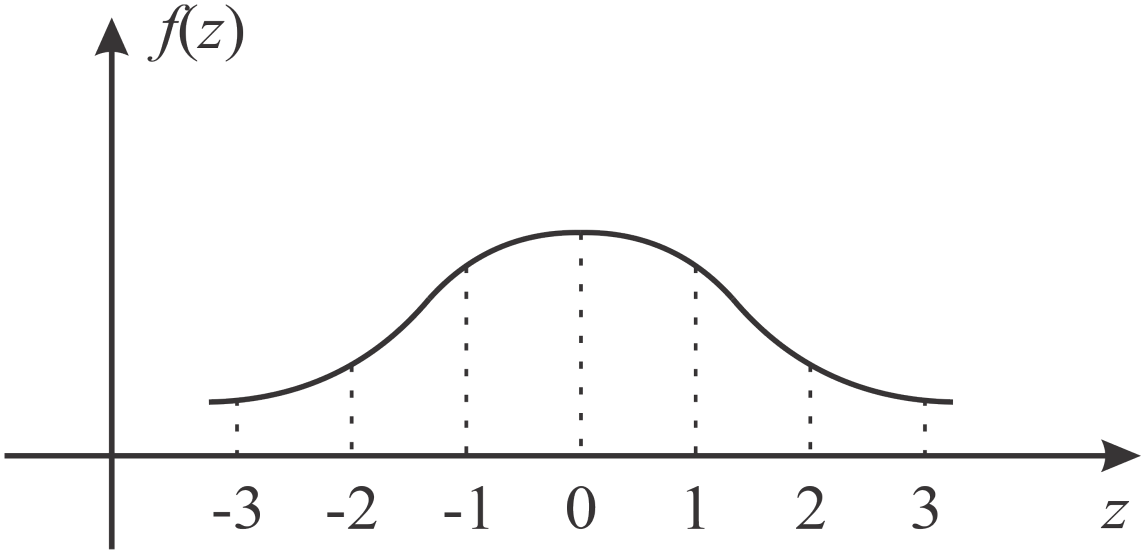
Gráfico 3 – Curva de Gauss



Quando um conjunto de dados está de acordo com uma distribuição normal, podemos obter as probabilidades desejadas com base no valor da média $\bar{X}$ e do desvio padrão σ.

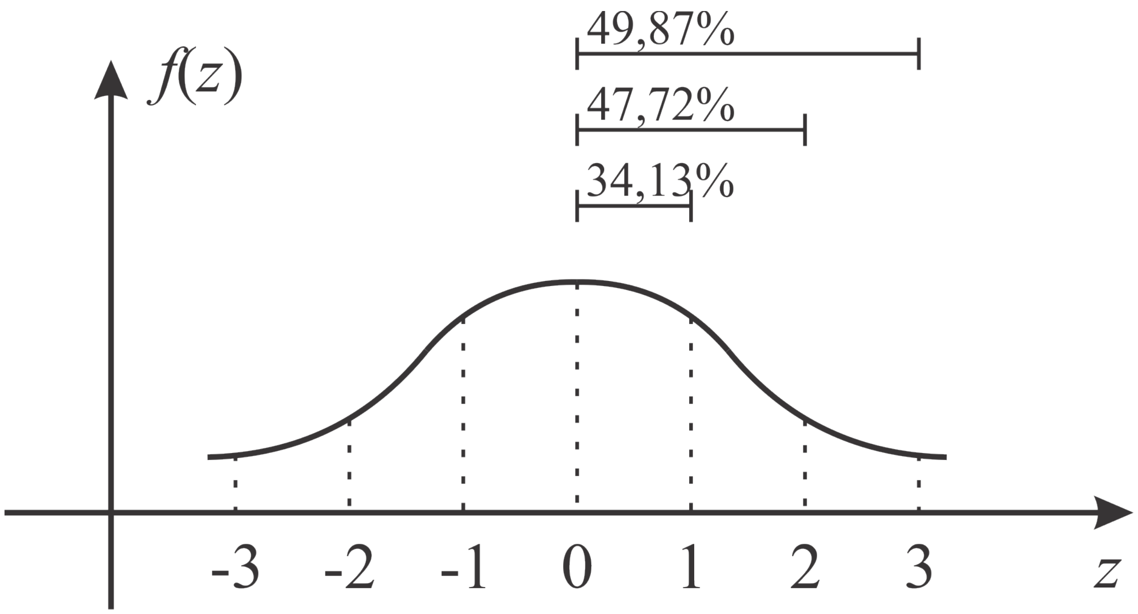
O cálculo da probabilidade de que a variável *X* esteja em determinado intervalo ocorre pela divisão do gráfico em três partes iguais localizadas à direita da média e em três partes iguais localizadas à esquerda da média. Cada uma dessas partes tem comprimento igual ao desvio padrão, e as probabilidades associadas a cada uma dessas partes são as mesmas, independentemente de qual problema é considerado, de qual valor está relacionado à média e de qual valor corresponde ao desvio padrão. Assim, é possível fazer uma parametrização para facilitar a resolução de problemas relacionados a uma distribuição normal. Consideramos um parâmetro *z*, em que cada unidade de *z* corresponde a uma unidade do desvio padrão relacionado ao problema a ser resolvido. O valor de *z* igual a 0 corresponde à média. Valores acima da média correspondem a valores positivos para *z*, e valores abaixo da média correspondem a valores negativos para *z*.

Gráfico 4 – Parametrização



Para cada porção do gráfico, temos uma porcentagem associada. Sabemos que da média para baixo temos 50% de probabilidade e da média para cima também temos uma probabilidade de 50%. Para as partições feitas, temos as probabilidades associadas a cada uma delas. Por exemplo, quando z = 1, a probabilidade corresponde a 34,13%; quando z = 2, a probabilidade é de 47,72%; e quando z = 3, a probabilidade corresponde a 49,87%.

Gráfico 5 – Probabilidades associadas às partições



Essas porcentagens foram obtidas por meio de cálculos específicos. Na resolução de problemas, podemos utilizar uma tabela de distribuição normal que contém os valores associados a diversos valores de *z*.

Tabela 2 – Tabela de distribuição normal

Distribuição normal										
Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	032257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3448	0,3461	0,3485	0,3508	0,3531	0,3534	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4396	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,44884	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,1591	0,4599	0,4608	0,4616	0,4625	0,433
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4703
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4820	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	034890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916

2,4	0,4918	0,4922	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Para relacionarmos *z* com os dados do problema e, consequentemente, para obtermos a probabilidade desejada, utilizamos a fórmula

$$z = \frac{X - \bar{X}}{\sigma}$$

em que é possível escrever os dados do problema em função de *z* e, em seguida, obter a respectiva probabilidade.

Exemplo:

Considere um projetor multimídia cuja lâmpada tem vida útil de 5000 horas com desvio padrão de 300 horas. Supondo que os dados estão de acordo com uma distribuição normal, determine a probabilidade de que um projetor, selecionado ao acaso, tenha lâmpada com vida útil entre 5000 e 5500 horas.

Como a média é igual a 5000 horas, desvio padrão é igual a 300 horas e *X* é igual a 5500 horas, podemos obter o respectivo valor de *z* fazendo:

$$z = \frac{X - \bar{X}}{\sigma}$$

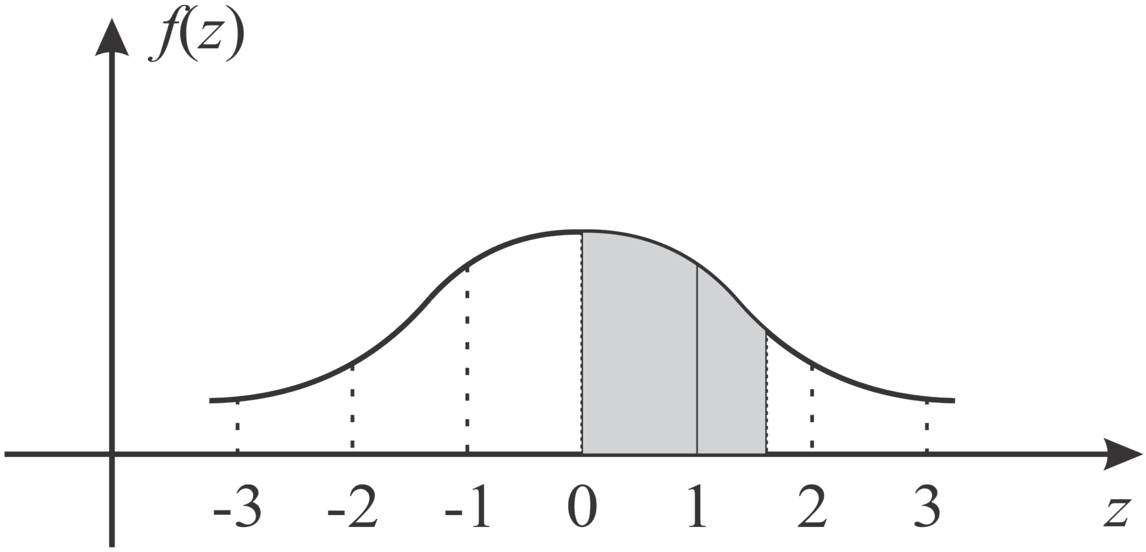
$$z = \frac{5500 - 5000}{300}$$

$$z = \frac{500}{300}$$

$$z = 1{,}67$$

Como z = 1,67, a região do gráfico que contém os dados que estão entre 5000 horas e 5500 horas é:

Gráfico 6 – Valor de z



A porcentagem que corresponde a essa região está na linha referente a 1,6 e na coluna referente a 0,07. Assim, a probabilidade quando z é igual a 1,67 corresponde a 0,4525, que na forma de porcentagem é igual a 45,25%:

Tabela 3 – Tabela de distribuição normal

Distribuição normal										
Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	032257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3448	0,3461	0,3485	0,3508	0,3531	0,3534	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830

1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4396	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,44884	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,1591	0,4599	0,4608	0,4616	0,4625	0,433
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4703

Logo, a probabilidade de que a lâmpada de um projetor selecionado ao acaso tenha vida útil entre 5000 e 5500 horas corresponde a 45,25%, ou seja, P (5000 ≤X ≤ 5500) = 45,25%.

Exemplo:

Considere um projetor multimídia cuja lâmpada tem vida útil de 5000 horas com desvio padrão de 300 horas. Supondo que os dados estão de acordo com uma distribuição normal, determine a probabilidade de que um projetor, selecionado ao acaso, tenha lâmpada com vida útil entre 4700 e 5600 horas.

Para calcularmos o valor de *z*, precisamos da diferença entre a variável *X* e a média. Como temos um intervalo em que há um valor menor do que a média e um valor maior do que a média, precisamos dividir o problema em duas partes, uma considerando a probabilidade no intervalo [4700, 5000] e outra considerando a probabilidade no intervalo [5000, 5600]. Após calcularmos essas duas probabilidades, somamos os resultados obtidos.

Para 4700 ≤ X ≤ 5000, temos:

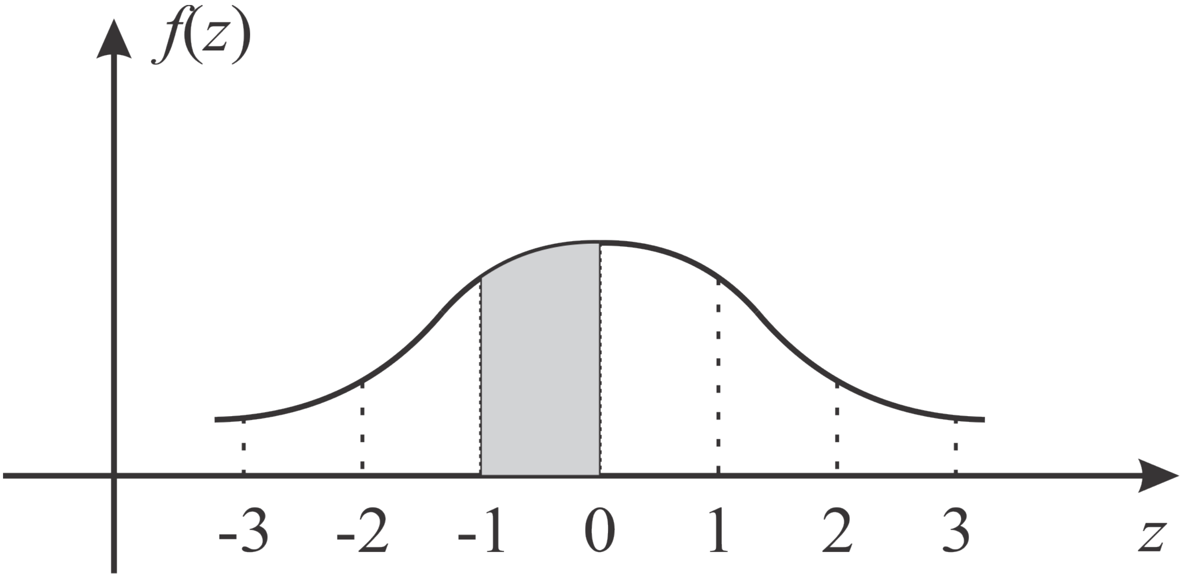
$$z = \frac{X - \bar{X}}{\sigma}$$

$$z = \frac{4700 - 5000}{300}$$

$$z = \frac{-300}{300}$$

$$z = -1$$

Gráfico 7 – Valor de z



Como o gráfico é simétrico em relação à média, z = -1 corresponde a z = 1.

Tabela 4 – Tabela de distribuição normal

Distribuição normal										
Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	032257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3448	0,3461	0,3485	0,3508	0,3531	0,3534	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015

Assim, P (4700 ≤ X ≤ 5000) = 34,13%.

Para 5000 ≤ X ≤ 5600, temos:

$$z = \frac{X - \bar{X}}{\sigma}$$

$$z = \frac{5600 - 5000}{300}$$

$$z = \frac{600}{300}$$

$$z = 2$$

Gráfico 8 – Valor de z

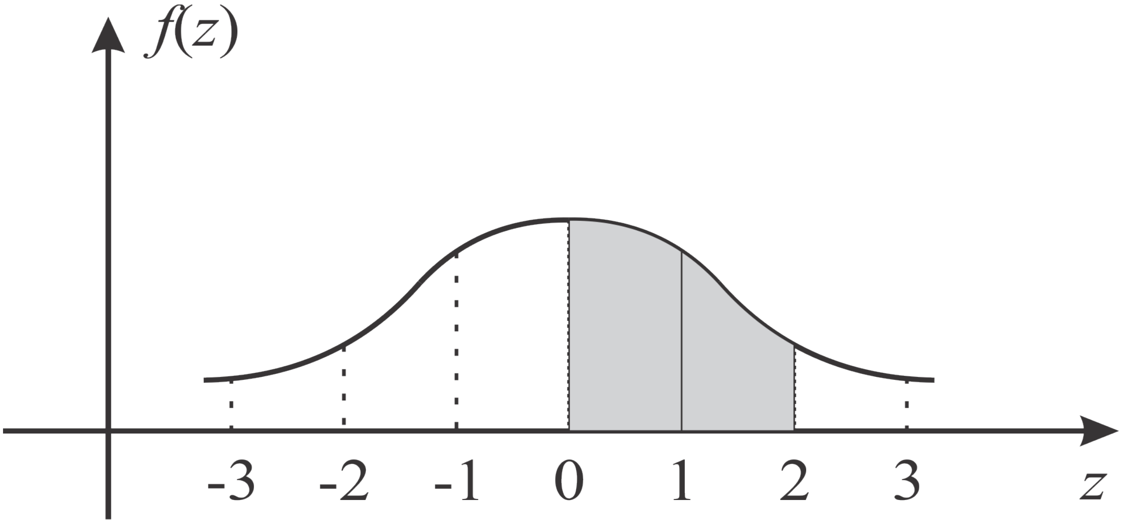


Tabela 5 – Tabela de distribuição normal

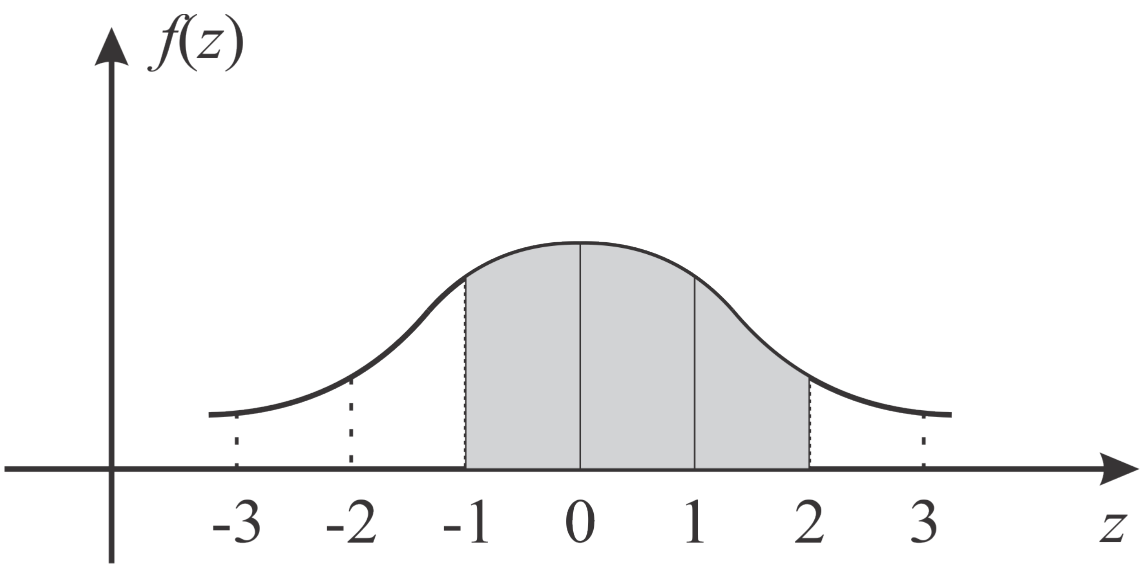
Distribuição normal										
Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	032257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549

0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3448	0,3461	0,3485	0,3508	0,3531	0,3534	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4396	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,44884	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,1591	0,4599	0,4608	0,4616	0,4625	0,433
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4703
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817

Logo, P (5000 ≤ X ≤ 5600) = 47,72%.

A probabilidade de ter vida útil entre 4700 e 5600 horas corresponde à região abaixo da média (z = -1) mais a região acima da média (z = 2):

Gráfico 9 – Valor de z



Portanto:

$$P\ (4700 \leq X \leq 5600) = P\ (4700 \leq X \leq 5000) + P\ (5000 \leq X \leq 5600)$$

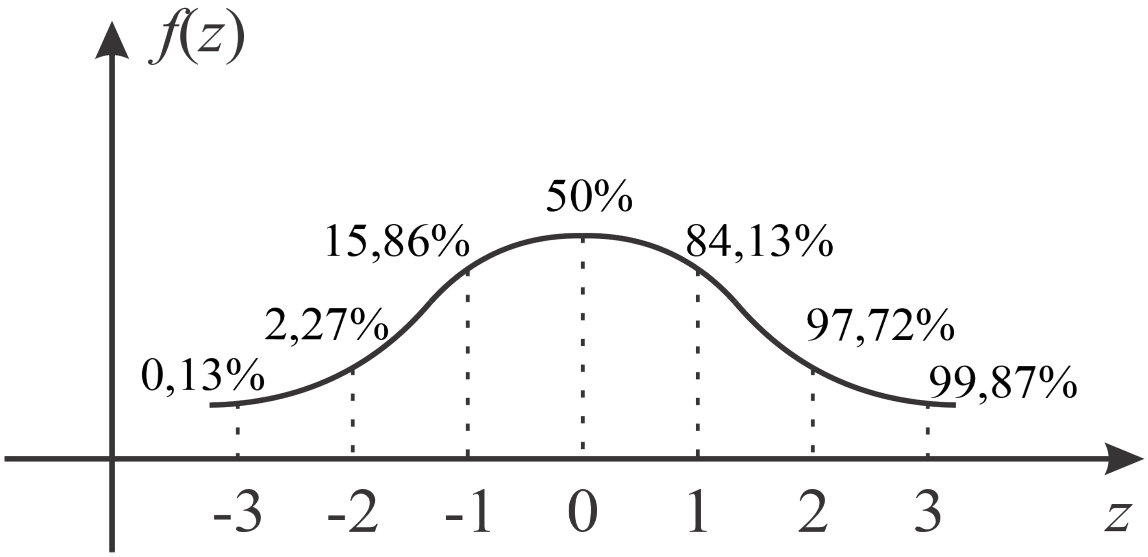
$$P\ (4700 \leq X \leq 5600) = 34{,}13\% + 47{,}72\%$$

$$P\ (4700 \leq X \leq 5600) = 81{,}85\%$$

A probabilidade de que a lâmpada de um projetor selecionado aleatoriamente tenha vida útil entre 4700 e 5600 horas é de 81,85%.

Podemos utilizar o Python para resolver problemas relacionados a uma distribuição normal por meio da função norm. Utilizaremos a opção cdf que está associada a uma distribuição normal cumulativa. Neste caso, quando z = 1, por exemplo, o resultado apresentado pelo Python considera a porção abaixo da média que corresponde a 50% mais a porção associada a z = 1 que é igual a 34,13%, totalizando 84,13%.

Gráfico 10 – Valor de z



Portanto, para valores de *z* maiores do que 0, calculamos a porcentagem acumulada e subtraímos 0,5, que corresponde a 50%, para termos apenas a porcentagem desejada. Para valores abaixo da média, basta fazermos 50% menos o resultado obtido.

Assim, para valores de *X* acima da média $\bar{X}$, podemos fazer:

Figura 9 – Comandos

```
import scipy.stats

media=

desvio_padrao=

X=

p=scipy.stats.norm(media,desvio_padrao).cdf(X)-0.5

print(p)
```

E para valores de *X* abaixo da média $\bar{X}$, podemos fazer:

Figura 10 – Comandos

```
import scipy.stats

media=

desvio_padrao=

X=

p=0.5-scipy.stats.norm(media,desvio_padrao).cdf(X)

print(p)
```

informando o valor da média, do desvio padrão e de *X*.

Exemplo:

Considere um projetor multimídia cuja lâmpada tem vida útil de 5000 horas com desvio padrão de 300 horas. Supondo que os dados estão de acordo com uma distribuição normal, determine por meio do Python a probabilidade de que um projetor, selecionado ao acaso, tenha lâmpada com vida útil entre 5000 e 5500 horas.

Utilizando a sequência de comandos

Figura 11 – Comandos

```
import scipy.stats

media=5000
```

```
desvio_padrao=300

X=5500

p=scipy.stats.norm(media,desvio_padrao).cdf(X)-0.5

print(p)
```

temos:

Figura 12 – Resultado dos comandos

```
0.4522096477271853
```

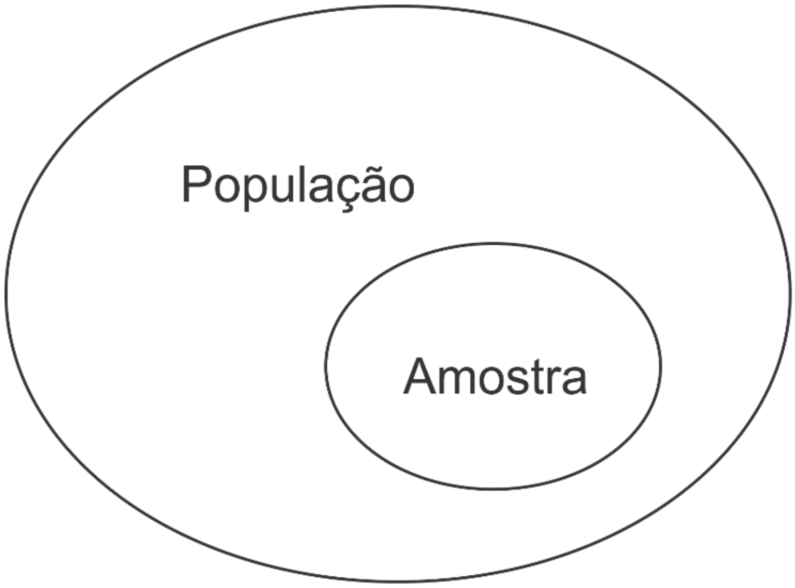
que corresponde a 45,22%. A diferença entre o valor obtido por meio do Python e o valor obtido por meio da tabela (45,25%) é que, no uso da tabela, foi feito um arredondamento no valor de *z*.

TEMA 4 – ESTATÍSTICA INDUTIVA (AMOSTRAGEM)

Sabemos que a estatística é uma área da matemática cujo objetivo é coletar, organizar, analisar e interpretar dados.

Os dados podem compor dois tipos de conjuntos, um chamado de *população* e outro chamado de *amostra*. A população está associada ao conjunto formado por todos os dados relacionados a um fenômeno. A amostra é um subconjunto da população.

Figura 13 – População e amostra



Quando não é possível realizar estudos com a população, utilizamos uma amostra. Em muitos casos, dados obtidos por meio de uma amostra são utilizados para inferências relacionadas a uma população. Para que esse processo tenha consistência, precisamos utilizar técnicas apropriadas para isso, tais como uma amostra significativa e aleatória em relação à população.

Sabendo o que é estatística, população e amostra, podemos pensar em como é realizada a amostragem, ou seja, a seleção de elementos para compor a amostra.

O primeiro passo é considerarmos uma amostra significativa para que o estudo seja confiável. Como não estamos considerando a população, mesmo que a amostra seja confiável, é comum que exista um erro entre os dados da amostra em relação aos dados da população. Na estatística inferencial, há técnicas de controle desses erros.

Uma amostra é considerada aleatória simples quando toda a amostra possível de mesmo tamanho tem as mesmas chances de ser selecionada. Uma forma de obtermos uma amostra aleatória simples é atribuirmos um número diferente para cada elemento da população e, com a geração de números aleatórios, selecionar os elementos que vão compor a amostra. É uma espécie de sorteio, no qual os números aleatórios podem ser obtidos de tabelas ou de aplicativos. Também é possível escolher os elementos que vão compor a amostra seguindo técnicas como amostra estratificada, amostra por agrupamento ou amostra sistemática.

Uma amostra estratificada é utilizada quando precisamos de elementos de diferentes segmentos da população. A população é dividida em grupos denominados de *estratos*, e as amostras são obtidas aleatoriamente de cada estrato. Assim, é possível garantir que cada um desses grupos terá uma amostra significativa relacionada.

No que se refere à amostra por agrupamento, a população naturalmente está dividida em grupos distribuídos geograficamente, cada um deles contendo características similares. Para obter uma amostra por agrupamento, a população é dividida em grupos, e todos os membros de um ou de mais grupos são selecionados. É importante cuidar para que não sejam selecionados os membros de todos esses grupos. Também é importante verificar se todos os grupos possuem características similares.

Considerando agora uma amostra sistemática, temos números associados a todos os membros da população que são ordenados seguindo determinado critério. Com base nesses números, escolhe-se um valor inicial, e os membros são selecionados de intervalos regulares. É uma amostragem fácil de ser utilizada.

No que se refere ao tamanho *n* da amostra, há diferentes formas de obter essa informação. Uma delas é a fórmula de Slovin, que leva em consideração o tamanho da população *N* e a margem de erro *e* aceitável.

$$n = \frac{N}{1 + Ne^2}$$

Essa fórmula, bastante simples, é utilizada quando não conhecemos o desvio padrão ou um nível de confiança associado ao estudo.

Exemplo:

Considerando que será realizado um estudo relacionado a 20.000 pessoas e que apenas uma algumas delas fará parte de forma efetiva, qual é o tamanho da amostra considerando uma margem de erro de 2%?

Para este problema, N = 20000 e e = 2% = 0,02. Utilizando a fórmula de Slovin, temos

$$n = \frac{N}{1 + Ne^2}$$

$$n = \frac{20000}{1 + 20000 \times 0{,}02^2}$$

$$n = \frac{20000}{1 + 20000 \times 0{,}0004}$$

$$n = \frac{20000}{1 + 8}$$

$$n = \frac{20000}{9}$$

$$n = 2222{,}22$$

Arredondando para o próximo inteiro:

$$n = 2223$$

Para esse estudo, é preciso considerar uma amostra de 2.223 pessoas para que a margem de erro seja de 2%.

Em Python, podemos fazer:

Figura 14 – Comandos

```
N=20000

e=0.02

n=N/(1+N*e**2)

print(f'Tamanho da amostra: {n}')
```

Figura 15 – Tamanho da amostra

```
Tamanho da amostra: 2222.222222222222
```

Para obtermos o resultado já com o arredondamento, podemos utilizar a função ceil da biblioteca math:

Figura 16 – Resultado com arredondamento

```
from math import ceil

N=20000

e=0.02

n=ceil(N/(1+N*e**2))

print(f'Tamanho da amostra: {n}')
```

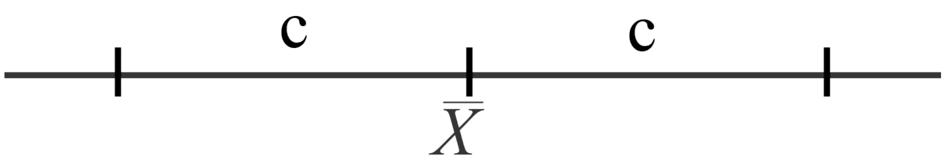
Figura 17 – Tamanho da amostra

```
Tamanho da amostra: 2223
```

TEMA 5 – ESTATÍSTICA INDUTIVA (CONFIANÇA)

O intervalo de confiança está associado à probabilidade de um valor desconhecido associado a um dado nível de confiança estar nesse intervalo. Para determinarmos o intervalo de confiança de uma variável, precisamos de uma determinada estimativa e de uma dada margem de erro. Por exemplo, podemos determinar um intervalo de confiança em relação à média.

Gráfico 11 – Intervalo de confiança



O termo *c* corresponde ao erro amostral e é calculado por meio da fórmula

$$c = Z\frac{\sigma}{\sqrt{n}}$$

em que

Z é a distribuição normal padronizada

σ é o desvio padrão da população

n é o tamanho da amostra

Tendo o valor de c, o intervalo de confiança é dado por

$$\bar{X} - c \leq \mu \leq \bar{X} + c$$

onde μ é a média associada à população.

Exemplo:

Em uma indústria de produtos alimentícios, o peso médio associado a pacotes de macarrão é de 504 gramas com desvio padrão de 10 gramas. Qual é o intervalo de confiança para um nível de confiança igual a 90% considerando uma amostra de 200 pacotes?

Como estamos considerando um nível de confiança igual a 90%, o primeiro passo é dividirmos 90% por 2, o que resulta em 45% = 0,45. Essa divisão é feita porque na tabela de distribuição normal temos os valores associados à metade da curva de Gauss. Por meio da tabela, sabemos que 0,45 corresponde a z = 1,65.

Tabela 6 – Tabela de distribuição normal

Distribuição normal										
Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	032257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3448	0,3461	0,3485	0,3508	0,3531	0,3534	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4396	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,44884	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,1591	0,4599	0,4608	0,4616	0,4625	0,433
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4703
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4820	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	034890

Para calcularmos *c*, temos:

$$\sigma = 10$$

$$n = 200$$

$$Z = 1{,}65$$

$$c = Z\frac{\sigma}{\sqrt{n}}$$

$$c = 1{,}65\frac{10}{\sqrt{200}}$$

$$c = 1{,}65\frac{10}{14{,}1421356237}$$

$$c = 1{,}65 \times 0{,}70710678119$$

$$c = 1{,}16672618896$$

Sabendo que c = 1,17 e que $\bar{X} = 504$, o intervalo de confiança é dado por

$$\bar{X} - c \leq \mu \leq \bar{X} + c$$

$$504 - 1{,}17 \leq \mu \leq 504 + 1{,}17$$

$$502{,}83 \leq \mu \leq 505{,}17$$

FINALIZANDO

Aprendemos a resolver problemas envolvendo média, moda e mediana, variância e desvio padrão de forma analítica e por meio do Python. Aprendemos os conceitos iniciais relacionados à probabilidade e o que é uma distribuição normal e resolvemos problemas práticos associados a este tema. Aprendemos também a determinar o tamanho de uma amostra utilizando a fórmula de Slovin e a obter um intervalo de confiança.